Orientation-dependent *Dxz4* contacts shape the 3D structure of the inactive X chromosome

G. Bonora[1*], X. Deng[2*], H. Fang[2], V. Ramani[1], R. Qiu[1], J. Berletch[2], G. N. Filippova[2], Z. Duan[3,4], J. Shendure[1], W.S. Noble[1,5], C.M. Disteche[2,6]

1) Genome Sciences, University of Washington, Seattle WA., USA; 2) Pathology, University of Washington, Seattle, WA., USA; 3) Institute for Stem Cell and Regenerative Medicine, University of Washington, Seattle WA., USA; 4) Hematology, University of Washington, Seattle WA., USA; 5) Computer Science and Engineering, University of Washington, Seattle WA., USA; 6) Medicine, University of Washington, Seattle WA., USA

*Equal contribution

Correspondence: C.M. Disteche, cdistech@uw.edu; W. S. Noble, william-noble@uw.edu

Abstract

The mammalian inactive X chromosome (Xi) condenses into a bipartite structure with two superdomains of frequent long-range contacts separated by a boundary or hinge region. Using in situ DNase Hi-C in mouse cells with deletions or inversions within the hinge we show that the conserved repeat locus *Dxz4* alone is sufficient to maintain the bipartite structure and that *Dxz4* orientation controls the distribution of long-range contacts on the Xi. Frequent long-range contacts between *Dxz4* and the telomeric superdomain are either lost after its deletion or shifted to the centromeric superdomain after its inversion. This massive reversal in contact distribution is consistent with the reversal of CTCF motif orientation at *Dxz4*. De-condensation of the Xi after *Dxz4* deletion is associated with partial restoration of TADs normally attenuated on the Xi, and with an increase in chromatin accessibility and CTCF binding, but few changes in gene expression, in accordance with multiple epigenetic mechanisms ensuring X silencing. We propose that *Dxz4* represents a structural platform for frequent long-range contacts with multiple loci in a direction dictated by the orientation of a bank of CTCF motifs at *Dxz4*, which may work as a ratchet to form the distinctive bipartite structure of the condensed Xi.

Mammalian X chromosome inactivation (XCI) results in the silencing of one of the two X chromosomes in female somatic cells, which is initiated by expression of the lncRNA *Xist* from the inactive X chromosome (Xi). This is followed by cis-spreading of *Xist* RNA and recruitment of specific proteins that implement repressive epigenetic changes, including loss of RNA polymerase II and of active histone marks such as H3-H4 acetylation and H3 methylation at lysine 4, as well as enrichment in silencing marks such as H3 tri-methylation at lysine 27 (H3K27me3) and H2A ubiquitination at lysine 119. These initial epigenetic modifications are followed by accumulation of the histone variant macroH2A, tri-methylation of H3 at lysine 9 and of H4 at lysine 20, together with DNA methylation at CpG islands[1,2]. Drastic changes in the structure of the Xi—most visibly condensation—take place once it becomes silenced. Early microscopy studies reported a condensed chromatin body (Barr body) specifically present in female cells, and later shown to represent the Xi, often located at the nuclear periphery or adjacent to the nucleolus[3-5].

Genome-wide chromosome conformation capture (Hi-C) studies in human and mouse cells and tissues demonstrate that condensation of the Xi is associated with the formation of a characteristic bipartite 3D structure[6-10]. Allele-specific Xi contact maps reveal two superdomains of condensation separated by a hinge. Long-range contacts are frequent within each superdomain, but are not observed between superdomains. While the genomic content of the superdomains differs between human and mouse, the hinge region is partially preserved and contains the macrosatellite repeat locus *DXZ4/Dxz4* in both species[6-10]. The *DXZ4/Dxz4* loci encode lncRNAs and bind CTCF and components of the ring-shaped cohesin complex only on the Xi, while the loci are methylated on the active X (Xa), preventing CTCF binding[11-15]. CTCF and cohesin are two of the main organizers of nuclear structure[16-19]. New studies have shown that highly dynamic chromatin loops are formed by progressive extrusion of the chromatin fiber due to the movement of extruding factors (EFs) such as cohesin rings. Extrusion proceeds until a boundary element (BE) bound to chromatin, such as CTCF, is encountered, which

stalls the loop and ultimately defines topologically associated domains (TADs)[20,21]. Within a TAD, loops can be continuously formed by loading of cohesin by the SCC2/SCC4 complex, processive loop extension, and release by WAPL[22,23]. Convergent CTCF binding motifs (i.e. facing each other) at the base of a loop favor strong interactions and the inversion of CTCF sites disrupts loop formation[10,24,25]. In the case of *DXZ4/Dxz4* CTCF motifs are arranged in tandem orientation at the locus, with an estimated 10-100 copies in human[13], and 14 copies in mouse[11], which was confirmed using Blast alignment of a 34bp conserved CTCF binding sequence. How the CTCF motif arrangement influences long-range chromatin contacts on the Xi is unknown. In addition to *Dxz4*, the mouse hinge region originally defined using Hi-C also contains the mouse-specific minisatellite repeat *Ds-TR* whose function is unknown[6,11]. Both *Dxz4* and *Ds-TR* loci bind nucleophosmin, an essential component of the nucleolus, and could represent a large nucleolus-associated domain (NAD) that may help position the Xi near the nucleolus[6,14].

To determine the role of each element of the hinge including *Dxz4* and *Ds-TR* in the maintenance of the 3D structure of the mouse Xi in relation to its silencing and nuclear positioning in somatic cells, we used allele-specific CRISPR/Cas9 editing to induce different sized deletions and inversions specifically targeted to the Xi. We then tested effects of these modifications on the overall 3D structure of the Xi using *in situ* DNase Hi-C[26]. High resolution allele-specific analyses were done to assess changes in the distribution of contacts and in the TAD structure along the X chromosomes. We scored these changes in relation to CTCF binding profiles obtained by ChIP-seq and to chromatin accessibility profiles obtained by ATAC-seq. Gene expression changes were measured by RNA-seq, considering genes normally silenced by XCI and genes that escape XCI. Finally, we determined the effects of genomic alterations of the hinge on the position of the Xi in the nucleus relative to the nuclear periphery and the nucleolus.

Results

1. The integrity of the superdomains on the mouse Xi depends on *Dxz4*, but not *Ds-TR*

To evaluate the role of specific elements located within the hinge that separates superdomains of long-range intrachromosomal interactions on the mouse Xi, we used allele-specific CRISPR/Cas9 to induce genomic alterations in Patski fibroblast cells, in which skewed XCI and frequent species-specific polymorphisms (1/93bp using validated Sanger MM10 *Mus spretus* SNPs) allowed us to distinguish the Xi from C57BL/6 (BL6) and the Xa from *Mus spretus*[6,27,28]. Xi-specific deletions or inversions were induced using pairs of small guide RNAs (sgRNAs) that flank either most of the hinge region or only some of its elements (Supplementary Table S1). We isolated two independent single-cell clones with a large 127kb deletion of the hinge that includes both *Dxz4* and *Ds-TR* (hereafter Del-hinge a and b) (Supplementary Fig. S1a). We also derived independent single-cell clones with either a 44kb deletion of *Dxz4* alone (hereafter Del-Dxz4), a 44kb inversion of *Dxz4* (hereafter Inv-Dxz4), a 37kb deletion of *Ds-TR* (hereafter Del-Ds-TR) alone, or a small 907bp inversion of 2 of 3 CTCF binding sites located at the 5'end of *Ds-TR* (hereafter Inv-5' Ds-TR) (Supplementary Fig. S1a). Sanger DNA sequencing analyses verified that the genomic alterations obtained by CRISPR/Cas9 editing were specific to the Xi allele, and Del-hinge clone a was also confirmed by FISH (Supplementary Fig. S1b, c and data not shown).

Next, in situ DNase Hi-C of the edited cell clones was performed in comparison to wild-type (WT) Patski cells using an established method (Supplementary Table S2)[6,26]. Contact maps for the Xi showed that a large deletion incorporating both *Dxz4* and *Ds-TR* (Del-hinge) as well as a deletion of *Dxz4* alone (Del-Dxz4) dramatically disrupted the bipartite structure of the Xi normally observed in wild-type cells (WT) (Fig. 1a,b). Inversion of *Dxz4* was associated with persistence of the Xi bipartite structure, but caused extensive re-distribution of contacts as described below. In contrast, deletion of *Ds-TR* (Del-Ds-TR) alone or alteration of its 5'end (Inv-5' Ds-TR) did not affect the Xi bipartite structure or the contact distribution. There was no apparent change in the contact maps of the Xa or of the autosomes in any of the deleted/inverted cell clones (Supplementary Fig. S2a). We conclude that *Dxz4* alone is necessary for the formation of the two superdomains on the Xi.

2. *Dxz4* contacts with loci in the telomeric superdomain are disrupted by *Dxz4* deletion or inversion

To increase the total number of allelic reads for better resolution, two pooled sets of Hi-C contacts were generated: a data set representing Del-hinge and Del-Dxz4 (designated Del-hinge/Dxz4) and another representing wild-type and Del-Ds-TR (designated WT*) (Supplementary Table S2). Allelic contact maps for each individual data set are very similar to those obtained from the pooled data, justifying the pooling (Fig. 1a; Supplementary Fig. S2a,b). Pearson correlation-transformed contact maps of the Xi and inferred 3D models clearly show that the hinge region, present in WT* almost disappears in Del-hinge/Dxz4 (Fig. 2a,b,d,e). In WT*, *Dxz4* is appears to be located at the telomeric edge of the hinge. In contrast, in Inv-Dxz4, a new de-condensed hinge forms telomeric to *Dxz4* as the locus is pulled into the centromeric superdomain, which becomes more condensed (Fig. 2a,c,d,f). In addition, the probability that *Firre*, another macrosatellite repeat locus that also binds CTCF on the Xi[14,29], comes in close proximity to *Dxz4* in 3D space is greater in Inv-*Dxz4* versus WT* or Del-hinge/Dxz4 (Fig. 2d-f). Zooming in to examine a 50Mb region around *Dxz4* clearly shows the loss of the separation between superdomains in Del-hinge/Dxz4, and the strong shift in contacts in Inv-Dxz4 (Fig. 2g-i). As expected no hinge is seen on the Xa (Supplementary Fig. S2c).

Hi-C analyses have shown that chromosomes are divided into two compartments, A and B, associated with open and closed chromatin, respectively[30]. To investigate whether disruptions of the Xi structure in Del-hinge/Dxz4 and Inv-Dxz4 result in changes to A/B compartment scores we decomposed WT*, Del-hinge/Dxz4, and Inv-Dxz4 allelic chromosomal contact maps into principal components. Principal component (PC) score profiles for autosomes and for the Xa were very similar between WT*, Del-hinge/Dxz4, and Inv-Dxz4, as expected, with PC1 capturing the A/B compartment structure (Fig. 2j-l; Supplementary Fig. S3a-c). In contrast, for the WT* Xi PC1, and to a lesser extent PC2 profiles capture the Xi bipartite structure rather than the underlying A/B compartment structure (Fig. 2j). Del-hinge/Dxz4 Xi PC1 score profile no longer shows the distinctive switch in sign at *Dxz4* seen in WT* reflecting the disruption to the bipartite structure (Fig. 2j). Indeed, WT* and Del-hinge/Dxz4

Xi PC1 scores show much lower correlation (ρ = 0.32) than seen for by the Xa (ρ = 0.95) and autosomes (ρ > 0.9) (Fig. 2l; Supplementary Fig. S3c). Although the distinctive bipartite profile seen in the WT* Xi PC1 score is still evident for Inv-Dxz4 Xi PC1, their correlation is intermediate (ρ = 0.76), reflecting the changes to Inv-Dxz4 Xi long-range contacts (Fig. 2j-l).

Differential contact maps show that the largest changes in contact frequency on the Xi in Del-hinge/Dxz4 or Inv-Dxz4 versus WT* were located in the vicinity of *Dxz4* and showed unidirectionality (Fig. 3a,b; Supplementary Fig. S3d-g). Contacts between the hinge region and loci telomeric to the locus (chrX:75-115Mb) decreased in both Del-hinge/Dxz4 and Inv-Dxz4, indicating that *Dxz4* normally contacts loci in the telomeric superdomain of the Xi. In Del-hinge/Dxz4 new long-range contacts appear between the two Xi superdomains, consistent with *Dxz4's* role of insulating the two superdomains (Fig. 3a; Supplementary Fig. S3d,e). In Inv-Dxz4, new contacts appear between *Dxz4* and multiple loci in the centromeric superdomain (chrX:5-75Mb; Fig. 3b; Supplementary Fig. S3f,g). Strikingly, a new contact domain appears between *Firre* and *Dxz4*, while the contact domain between *Dxz4* and *Xist* present in WT* is lost (Fig. 3b; Supplementary Fig. S3f,g). These contact changes most probably result from the reversal of the orientation of CTCF motifs previously identified at *Dxz4,* which are mainly oriented toward the telomeric end of the mouse X in wild-type (Fig. 1b)[11].

The unidirectional nature of interactions between *Dxz4* and other Xi loci is clearly evident in virtual 4C plots based on Hi-C data using *Dxz4* as a viewpoint (Fig. 3c). Additional 4C plots generated for additional viewpoints along the Xi show that interactions between *Dxz4* and regions located as far as 40Mb telomeric to the locus (chrX:75-115Mb) are strongly reduced in Del-hinge/Dxz4 or Inv-Dxz4 versus WT* (Supplementary Fig. S3h). Note that along this 40Mb region on the WT* Xi the number of contacts with *Dxz4* varies with evidence for contact hotspots, for example between *Dxz4* and *Xist* (Fig. 3c; Supplementary Fig. S3h). Contact hotspots were also seen in Inv-Dxz4, especially between *Dxz4* and *Firre* (Fig. 3c).

To quantify changes in contact frequency extending across the hinge region at all scales, including very long-range contacts, we used a modified version of the coverage score measure[31] (see

Methods). The strong dip in coverage score due to lack of contacts between superdomains on the Xi in WT* was not seen in Del-hinge/Dxz4, but was mostly retained in Inv-Dxz4 (Fig. 3d). This analysis confirms extensive loss of long-range contacts between *Dxz4* and regions telomeric to the locus (chrX:75-115Mb) in Del-hinge/Dxz4, and gain of contacts between the two superdomains resulting in a higher coverage score at chrX:50-75Mb, while few changes occurred in distal regions (chrX:5-50Mb, 140-165Mb) (Fig. 3d). In the Inv-Dxz4 Xi, a dramatic decrease in the coverage score between chrX:75-110Mb and a corresponding increase between chrX:45-75Mb can be attributed to the reversal of *Dxz4* anchored contacts. Similar results were obtained when considering all cell lines: changes to the coverage score profile are only seen for the deletions that include *Dxz4* and the inversion of *Dxz4* (Supplementary Fig. S4a). Coverage scores for the Xa were similar between cell lines (Supplementary Fig. S4b). Violin plots showed a smaller range in negative coverage scores for Del-hinge and Del-Dxz4, and to a lesser extent for Inv-Dxz4, than for WT, Del-Ds-TR, and Inv-5' Ds-TR, confirming loss of very long-range contacts (Supplementary Fig. S4c). Hierarchical clustering using both Euclidean and Pearson correlation distance measures resulted in segregation of Xa's from Xi's, and within the Xi's, segregation of those in Del-hinge and Del-Dxz4 from the others (Supplementary Fig. S4d,e).

Intriguingly, the differential contact map between Inv-Dxz4 and WT* shows that a region immediately telomeric to *Dxz4* makes many new contacts with centromeric loci in Inv-Dxz4 Xi (red band runnings from *Dxz4* to centromere) (Fig. 3b). Interactions with telomeric loci are simultaneously lost (blue band from *Dxz4* to telomere). This represents a shift in the transition zone between the two superdomains. Examining this more closely using coverage scores based on higher resolution data (40kb), we found that the region of minimal interactions between the two superdomains shifts to the opposite side of *Dzx4* in Inv-Dxz4 Xi compared to WT*, while almost doubling in size (Fig. 3e). This shift in the superdomain boundary confirms the apparent repositioning of *Dxz4* towards the centromeric superdomain and away from the telomeric superdomain seen by comparing the 3D model of the Inv-Dxz4 Xi to that for the WT* Xi (Fig. 2d,f).

Taken together, these results demonstrate that *Dxz4* specifically contacts a number of regions

located telomeric to the locus in the wild-type Xi, and that deletion or inversion of *Dxz4* exerts their strongest effects in the vicinity of the locus. The unidirectional nature of contacts due to the orientation of CTCF binding sites is clearly demonstrated by contact changes observed after *Dxz4* inversion.

3. TADs on the Xi are partially restored after deletion or inversion of *Dxz4*

We and others have reported that the mouse Xi does not display prominent TADs, in contrast to the Xa or the autosomes[6,8,9,18]. To assess potential changes in short-range contacts around loci normally subject to XCI (e.g. *Eda2r*, *Zfx*), we generated high-resolution (40kb) Hi-C contact maps within 4Mb regions centered at these loci, which show that TADs, normally attenuated on the Xi, is partially restored on the deleted Xi in Del-hinge/Dxz4 and assumes a pattern more similar to that of the Xa (Fig. 4a). In contrast, loci that escape XCI such as *Ddx3x* show little change.

To quantify TAD number and distribution, we called TADs using a customized implementation of the insulation score[32]. Overall, a similar number of TADs were identified on the Xi and Xa, although TADs were attenuated on the Xi (Supplementary Table S3). Comparisons of standardized insulation score profiles for the entire Xi showed that insulation scores for Del-hinge/Dxz4 and Inv-Dxz4 differed from WT*. Short to medium range contacts increased in a 40Mb region telomeric to *Dxz4*, consistent with restoration of TADs in that region (Fig. 4b). Note that a particularly large decrease in insulation scores occurred in a region between chrX:63.25 and 66.75Mb (at 500kb). Zooming in on a 10Mb region around *Dxz4,* insulation scores appear quite different between the WT* Xi and the Xa, while the Del-hinge/Dxz4 Xi is more similar to the Xa (Fig. 4c). These phenomena were robust and observed at both 500kb and 40kb resolution (using a 3.5Mb and a 520kb sliding window, respectively), while profiles for the Xa were very similar between cell types (Supplementary Fig. S4f,g). These trends are reflected in both their distributions and hierarchical clustering (Fig. 4d,e; Supplementary Fig. S4h,i). Furthermore, Del-hinge/Dxz4 and WT* TADs called at 40kb resolution along the Xi show less overlap than do TADs called along the Xa (Supplementary Fig. S4j). We conclude that the TAD structure of the *Dxz4*-deleted Xi and to a lesser extent that of the Inv-Dxz4 Xi is intermediate between that of the wild-type Xi and Xa,

in line with an increase in short-range contacts on the deleted Xi.

4. CTCF binding to the Xi increases after deletion of the hinge

CTCF plays an important role in defining boundaries between contact domains, such as TADs[17-19]. We derived CTCF binding profiles by ChIP-seq for WT and Del-hinge Patski cells. ChIP-PCR analysis confirmed loss of CTCF binding at the deleted *Dxz4* locus in Del-hinge cells, while CTCF binding at the control *H19* gene was maintained (Supplementary Fig. S5a). For each CTCF peak with sufficient SNP coverage, an allelic proportion of SNP read coverage (*spretus*/(*spretus*+BL6)) was calculated. Average allelic proportions (Xa(Xa+Xi)) of X-linked CTCF peaks are markedly different between WT (~80%) and Del-hinge (~60%) (Supplementary Fig. S5g), with their distribution showing a pronounced shift toward lower values in the Del-hinge Xi compared to the wild-type Xi (Fig. 5a). This shift represents a substantial increase in CTCF binding along the Del-hinge Xi relative to the WT Xi, something not observed for autosomes, which exhibit overall allelic proportions of just under 50% (Fig. 5a; Supplementary Fig. S5f). Plots of CTCF peak d-scores (*spretus*/(*spretus*+BL6) – 0.5) confirm the increase in CTCF binding along the entire Xi specifically in Del-hinge (Supplementary Fig. S6a; S7a). Imprinting control regions near *H19* (maternally expressed gene) and within *Peg3* (paternally expressed gene) on chromosome 7 showed CTCF peaks only on the maternal and paternal allele, respectively (Supplementary Fig. S7b).

Diploid CTCF peaks with sufficient SNP coverage (5x) were designated as either *spretus*-specific or BL6-specific based on an allelic proportion (*spretus*/(*spretus*+BL6)) greater than 70% or lower than 30%, respectively, with remaining peaks classified as biallelic (Fig. 5b). The majority of X-linked CTCF peaks in Del-hinge are biallelic (>58%), whereas in WT there is a comparatively lower proportion of biallelic peaks (37%) with the majority being Xa-specific (>58%). This represents a pronounced increase in CTCF binding along the Xi after deletion of the hinge (Fig. 5b). By comparison, the majority of autosomal peaks are biallelic in both WT and Del-hinge (65-70%).

Interestingly, some CTCF peaks that appear on the deleted Xi occur near genes whose TAD

structure and expression change in Del-hinge cells. For example, at *Zfx,* a gene normally subject to XCI, new CTCF binding peaks (marked by arrows) located both within the gene and downstream of the gene appear on the Del-hinge Xi, consistent with more prominent TADs as well as reactivation of the gene (see below) (Fig. 4a; 5c). In contrast, genes that escape X inactivation both in WT and Del-hinge cells, for example *Eif2s3x*, show no change in CTCF peaks (Fig. 5c). The *Firre* locus, which strongly binds CTCF and nucleophosmin on the Xi[14,28], shows no change in CTCF binding peaks in Del-hinge cells (Supplementary Fig. S6b). We conclude that deletion of the hinge results in a general increase in CTCF binding along the Xi.

5. Chromatin accessibility increases on the Xi after deletion of the hinge

Chromatin accessibility was measured using ATAC-seq. As was done for CTCF, for both WT and Del-hinge ATAC peaks were called jointly across both alleles using unsegregated reads (Supplementary Table S5). The number of SNP-containing peaks was similar between WT and Del-hinge, and the vast majority of SNP-containing ATAC peaks were covered by at least 5 reads (Supplementary Fig. S8c,d). Although the overall autosomal allelic proportions (*spretus*/(*spretus*+BL6)) are similar for WT and Del-hinge (~50%), the X-chromosome allelic proportion for Del-hinge is much lower compared to WT (~70% compared to ~85%), indicative of an increase in ATAC peaks and chromatin accessibility (Supplementary Fig. S8f). Indeed, the WT and Del-hinge distributions of allelic proportion for ATAC peaks are very different for the X-chromosome (Fig. 5e; Supplementary Fig. S8g), while for autosomes the distributions overlap almost perfectly. Plots of ATAC peak d-scores (*spretus*/(*spretus*+BL6) – 0.5) show that the increase in chromatin accessibility occurs across the entire Xi in Del-hinge, rather than being restricted to regions around the *Dxz4* locus (Supplementary Fig. S6a). Autosomes, on the other hand, show little to no difference in d-score between WT and Del-hinge (Fig. 5e; Supplementary Fig. S7a).

Diploid ATAC peaks with sufficient SNP coverage (5x) were designated as *spretus*-specific or BL6-specific if they showed an allelic proportion (*spretus*/(*spretus*+BL6)) greater than 70% or lower

than 30%, respectively, and biallelic otherwise (Fig. 5f). In Del-hinge, a far smaller proportion of ATAC peaks are Xa-specific in Del-hinge (~55%) compared to WT (~90%), with a far higher proportion of biallelic peaks in Del-hinge (40%) compared to WT (~90%) (Fig. 5f). In contrast, autosomal ATAC peaks show comparable proportions of biallelic and allele-specific peaks, implying similar accessibility (Fig. 5f). Genome browser tracks clearly show a strong increase in ATAC peaks along the Xi (BL6 X-chromosome) in Del-hinge cells, something not seen along BL6 autosomes (Supplementary Fig. S6,S7). Zooming in on individual genes, the appearance of ATAC peaks around X-linked genes that are normally inactivated and become expressed from the Xi in Del-hinge cells, for example *Zfx*, is evident, while ATAC peaks are not altered around genes that escape XCI such as *Eif2s3x* (Fig. 5g). Taken together, results of the ATAC-seq analysis show a strong increase in Xi chromatin accessibility after deletion of the hinge, consistent with a role for *Dxz4* in maintaining condensation of the Xi. Furthermore, biallelic CTCF peaks and ATAC peaks showed spatial correlation in terms of their relative distance[33] on the Xa and Xi in both WT and Del-hinge (data not shown).

6. Deletion of the hinge has no effect on Xi positioning

We previously reported that the *Dxz4* and *Firre* loci on the Xi, but not on the Xa, are often preferentially located near the edge of the nucleolus, and that knockdown of *Firre* lncRNA caused a significant reduction (29%) in both the frequency of this preferred location and in the level of H3K27me3 on the Xi, visible as a 76% loss of nuclei with a cluster of H3K27me3 immunostaining[14]. Our current study shows that H3K27me3 immunostaining results with a strong Xi-staining cluster in ~90% of the nuclei from Del-hinge and Inv-Dxz4 cells (Fig. 6a and data not shown), suggesting that *Dxz4* deletion or inversion did not cause major changes in H3K27me3 enrichment on the Xi. To determine whether *Dxz4* deletion or inversion affects Xi positioning within the nuclei of Patski cells, we performed dual immunostaining for H3K27me3, to mark the Xi, and for nucleophosmin, to label the nucleolus. No significant changes in the frequency of Xi positioning either near the nuclear periphery or near the nucleolus were found (Fig. 6b). Thus, neither *Dxz4* deletion, nor its inversion, affect the position of the Xi in the nucleus of Patski cells.

7. Gene expression after modifications to the hinge/*Dxz4*

RNA-seq was performed in WT cells, in Del-hinge cell clones a and b, and in single clones for Del-Dxz4, Inv-Dxz4 and Del-Ds-TR. An allele-specific data analysis pipeline (see Methods) identified a similar set of genes that escape XCI in WT cells as in our previous study (29 escape genes; Supplementary Table S7)[34]. We initially focused our analyses on Del-hinge clone a, which is near-diploid, like WT cells. There was some evidence of reactivation of genes on the Xi, with 16 genes showing significant increased expression relative to WT, and no gene showing significant downregulation (log2 fold change >= 0.5 and adjusted p-value <= 0.05) (Fig. 7a,b; Supplementary Table S8). This unidirectional trend was not observed for the Xa, for which 55 genes showed a significant change in expression between Del-hinge and WT but with the change occurring in both directions (Fig. 7a,b and Supplementary Table S8). Interestingly, there was no correlation between the position of *Dxz4* and the location of Xi-alleles with expression changes (Fig. 7c). When X-linked genes were grouped in terms of their silencing/escape status in WT cells, we found that deletion of the hinge affected Xi-expression of more genes subject to XCI (12/209 or 6% of genes) than escape genes (4/29 genes or 14% of genes), although the proportion of changed escape genes was higher (Fig. 7c). In addition, some of the reactivated genes on the Xi in Del-hinge clone a (e.g. *Zfx*, *Eda2r*) showed a TAD structure more similar to that of the Xa as well as an increase in CTCF peaks and ATAC peaks (Fig. 4a; 5c,g).

Next, we compared Xi- and Xa-specific levels in additional cell lines including Del-hinge clone b Del-Dxz4, and Del-Ds-TR to levels in WT (Supplementary Fig. S9). As many as 77-85% of genes showed no or very low Xi-specific expression (Xi-TPM <0.2), indicating that silencing of most X-linked genes is largely maintained in these cell lines. There was a minor reactivation of the Xi in Del-hinge clone b and in Del-Dxz4, with the upregulated genes partially overlapping those observed in Del-hinge clone a, indicating some similarity but also variability between cell clones (Supplementary Fig. S9).

RNA-seq performed in two Inv-Dxz4 clones (a and b) and a corresponding subclone of WT cells

used to derive the inversion lines (Patski2-4 a and b, respectively) showed no evidence of reactivation of genes on the Xi (log2 fold change >= 0.5 and adjusted p-value <= 0.05) (Supplementary Fig. S10a, Table S9). Only two genes (*Msn* and *Clcn5*) showed a significant increase Xi-specific expression relative to WT, with another two genes (*Shroom2* and *Hnrnph2*) showing a significant decrease in expression. By RT-PCR, *Dxz4* expression was similar in Del-hinge, Inv-Dxz4 and WT (data not shown), but we could not separate expression from the Xa or Xi, due to lack of informative SNPs.

XCI is associated with multiple layers of epigenetic modifications, including DNA methylation at CpG islands of silenced genes[1,2]. To test for potential synergistic effects between condensation of the Xi and DNA methylation, we examined the effects of global DNA demethylation on X-linked gene expression in Del-hinge clone a versus WT cells. Consistent with the inhibitory role of DNA methylation on gene expression, RNA-seq analyses showed global upregulation of both autosomal and X-linked genes in WT and Del-hinge cells after 5-aza-2dC treatment (Supplementary Fig. S10b-d). In WT cells, 2328 autosomal and 129 X-linked genes were significantly upregulated, with only 640 autosomal and 27 X-linked genes being downregulated (log2 fold change >0.5, adjusted p-value <0.05) (Supplementary Fig. S10b-d). Similarly, in Del-hinge cells, 1318 autosomal and 75 X-linked genes were significantly upregulated and only 133 autosomal genes and no X-linked gene, downregulated. Demethylation caused an increase in the number of reactivated X-linked genes, which was similar in Del-hinge versus WT (Supplementary Fig. S10b-d and Supplementary Tables S10 and S11). Thus, deletion of the hinge with or without DNA demethylation is insufficient to cause massive reactivation of genes on the Xi, despite the observed de-condensation and increase in chromatin accessibility.

Discussion

Our study shows that *Dxz4* is the sole element necessary for maintenance of the condensed 3D bipartite configuration of the inactive X chromosome. Two previous studies based on large deletions (200-300kb) obtained in mouse ES cells or in human cells also reported disruption of the Xi bipartite structure[7,8]. However, we find that a much smaller deletion (44kb) of *Dxz4* alone is sufficient to cause

massive de-condensation of the Xi. Remarkably, the bipartite structure persists after inversion of the *Dxz4* locus, indicating that it plays an insulating role in either orientation albeit we observed a massive redistribution of long-range contacts from the telomeric to centromeric superdomain. This is presumably due to a reversal of the mostly unidirectional CTCF motifs at the *Dxz4* locus causing them to make a series of new contacts with loci in the opposite direction along the Xi.

We demonstrate that after deletion of *Dxz4*, the TAD structure at individual loci on the Xi is restored to a structure more similar to that of the Xa and that there is an increase in CTCF binding and in chromatin accessibility measured by ATAC-seq, consistent with decondensation of the Xi. However, we observed little reactivation of X-linked genes that are normally silenced, even in the presence of a demethylating agent, attesting to the multiple layers of controls that ensure silencing of the X, and suggesting that de-condensation and increased chromatin accessibility are not sufficient to elicit widespread reactivation on the Xi[1,36,37]. The limited Xi reactivation we observed was variable between cell clones, suggesting that decondensation of the Xi due to *Dxz4* deletion may facilitate random variable dysregulation of X-linked genes. Evidence of reduced expression of genes that escape XCI was reported in mouse ES cells with a hinge deletion[8], but was not found in our analysis of Patski fibroblasts. We also did not observe changes in H3K27me3 immunostaining, but we cannot exclude local changes around *Dxz4*. Localized loss of H3K27me3 and gain of H3K9me3 enrichment, as well as a change in Xi replication timing status, which shifted from early to late, were observed in the study of human cells with a 300kb deletion including *DXZ4*[7]. However, similar to our findings, there was little change with respect to gene expression in human cells. We found that the preferred locations of the Xi near the lamina or the nucleolus were apparently not disrupted by deletion or inversion of *Dxz4*. This was surprising since we previously reported that both *Dxz4* and *Ds-TR* bind nucleophosmin, and that *Dxz4* and *Firre* are often located near the nucleolus when on the Xi[14]. It is possible that the preferential nuclear location of the Xi may be facilitated by multiple loci including *Firre*, which is also bound by CTCF and nucleophosmin specifically on the Xi[14,29].

Our findings that an inversion of *Dxz4*, which includes its promoter, causes massive contact re-

distribution, suggest that the locus itself rather than its lncRNA is critical for the maintenance of contacts. This would not exclude *Dxz4* transcription as a potentially important factor, possibly by rendering the locus accessible, but this remains to be clarified. The human *DXZ4* and mouse *Dxz4* lncRNAs are expressed from both Xa and Xi alleles[13,14]. However, there are differences in the levels and types of transcripts produced from each allele: in human, the *DXZ4* transcripts include a long sense-transcript and short antisense-transcripts from both alleles, and a long antisense-transcript from the Xi only[13]. In mouse, a long sense transcript was detected from the Xa[11] and we reported higher expression in female versus male tissues and cell lines, consistent with Xi expression[14]. The small RNAs produced from the Xi at the human *DXZ4* locus may help recruit Argonaute proteins to facilitate DNA methylation[38]. DNA methylation at *DXZ4/Dxz4* specifically marks the active X, which would prevent CTCF binding[11,12,15].

Analyses of contact distributions show that the greatest level of disruption of the bipartite structure of the Xi after *Dxz4* deletion or inversion is in the vicinity of *Dxz4*, which is in agreement with a study in human where a hinge deletion coincides with the disruption of an interaction compartment (as reflected by principal eigenvectors) in the vicinity of *DXZ4*[7]. A novel finding in our study is that *Dxz4* mainly contacts loci located in the telomeric superdomain and that these long-range contacts are either lost after *Dxz4* deletion or shifted to the centromeric superdomain after *Dxz4* inversion. CTCF strongly binds at *Dxz4* on the Xi via unidirectional conserved CTCF binding motifs embedded in the tandem repeats of the macrosatellite locus[11,14]. The orientation of CTCF motifs is critical in the formation of chromatin loops by extrusion[10,20,21,24,39]. In the extrusion model, a cohesin ring complex (or other EF) is initially loaded onto two adjacent regions of chromatin in cis and then slides over the chromatin fiber in opposing directions, bringing increasingly remote regions together and resulting in the extrusion of a loop. The processivity of EFs is believed to be curtailed by BEs such as CTCF, apparently in a directional manner, with binding to convergent motifs being far more effective[10,20,21,24,39]. In this scenario, the particularly large bank of Xi-specific CTCF binding at *Dxz4*[11,14] would present a formidable barrier to EFs, halting chromatin fiber translocation at that end of the loop. At the other end of the loop,

EFs would remain relatively free to continue to extrude chromatin, unless they were to encounter another loop or another CTCF binding site preferably in opposite orientation (Fig. 7d). Thus, EFs would tend to stall at the *Dxz4* locus, impeded from moving through to the other side of the locus and effectively insulating the two superdomains from one another. This is consistent with polymer simulations highlighting the role of EFs in imposing insulation over large spatial and genomic distances simply by their translocation processivity being regulated by BEs such as CTCF[20]. Following this reasoning, regions immediately to the rear of the *Dxz4* locus (with directionality dictated by the orientation of the CTCF motifs) would be subject to reduced translocation, due to the EFs being stalled within the bank of CTCF sites. The diminished extrusion would result in fewer long-range contacts and reduced loop formation with chromatin taking on a more unraveled appearance, as we have shown in 3D models of the WT and Inv-Dxz4 Xi (Fig. 2d-f). This unraveling 'behind' *Dxz4* may also be a result of increased tension in the chromatin fiber due to the locus being pulled closer to *Xist* and *Firre* in the WT and Inv-Dxz4 Xi respectively.

Among the new long-range contacts found after *Dxz4* inversion, the most striking is a strong interaction between *Dxz4* and *Firre*, most likely mediated by the banks of CTCF binding sites at both loci[11,14,29]. While the orientation of CTCF motifs is not completely defined at *Firre* due to the repetitive nature of the locus motif searches indicate that a similar number of CTCF binding sites exist in either orientation. Any convergent CTCF motifs may pair with those at the inverted *Dxz4* locus, facilitating strong contacts. The new long-range contacts between *Dxz4* and other loci in Inv-Dxz4 do not normally form on the WT Xi, indicating that *Dxz4* contacts have little specificity, consistent with the lack of conservation of the *Dxz4* repeat units, except for the CTCF motifs and a short sequence (13bp) that could potentially represent a binding site for another BE[11]. The bank of CTCF binding sites at the inverted *Dxz4* locus apparently still has the ability to curtail the movement of cohesin, as described above for WT, resulting in a bunching of chromatin fibers and strong contact domain.

The tandem arrangement of repeats at *Dxz4* suggests that the locus is uniquely suited to participating in the condensation of the Xi telomeric superdomain by promoting the formation of a

multitude of loops. This leads us to propose a model in which *Dxz4* acts as a row of unidirectional obstacles that hook loops of chromatin for condensation (Fig. 7d). This model also suggests a possible mechanism by which the Xi may fold at the hinge but this remains to be further investigated[6]. The platform structure may behave in manner more akin to a ratchet than to velcro, since the locus orientation is important and translocation of chromatin is impeded unidirectionally. In such a scenario cohesin rings would sit along the *Dxz4* locus at sites paired with convergent CTCF sites with chromatin loops threaded through them. Indeed, we and others have reported a strong visible CTCF accumulation within the Xi in interphase nuclei, which co-localizes with *DXZ4/Dxz4*[40,41]. Although Xi-specific *Dxz4* contacts are represented as occurring simultaneously along the same chromatin fiber in the model and could be interpreted to be stable structures, in reality this might not be the case. This may well be a dynamic process, with contacts between distant X-linked loci and the *Dxz4* platform being transient albeit relatively frequent. And whether some or all units of the Xi tandem repeat are engaged in loop formation in a single cell is unclear. Single-cell analyses may help better understand cell-to-cell variability in the Xi contact distribution.

Methods

Allele-specific CRISPR/Cas9 editing of Patski cells

Patski cells are fibroblasts derived from 18dpc embryonic kidney from a cross between a BL6 female mouse with an *Hprt*^{BM3} mutation[42] and *Mus spretus* male. The cells were selected in HAT media such that the BL6 X chromosome is always inactive as verified in previous studies[27,28]. For allele-specific CRISPR/Cas9 editing sgRNAs designed using CHOPCHOP[43,44] were selected to include BL6 SNPs at the PAM site if available (Supplementary Fig. 1a and Table S1). Patski cells were transfected using Ultracruz transfection reagents (Santa Cruz). Verifications of the deletions/inversion were done using PCR together with Sanger sequencing to verify specific loss of the BL6 allele with the deletion and to verify junction sequences containing BL6 SNPs (Supplementary Fig. 1b and Table S2). Note that two

independent clones (a, b) with a deletion of the hinge were derived, while single clones were derived for other deletions or inversions. Del-hinge clone a, which has a near-diploid karyotype, was used for all studies, except for gene expression analyses in which we considered both clones a and b.

Immunostaining and FISH

Immunostaining of paraformaldehyde fixed nuclei using antibodies for H3K27me3 (Upstate/Millipore) and for nucleophosmin (Abcam) was done as described[14]. Nuclei were examined by fluorescence microscopy to score the position of the Xi marked by H3K27me3 with respect to the nuclear periphery and the edge of the nucleolus. A total of ~200 nuclei were scored by at least two different observers. DNA-FISH was done on metaphase and interphase cells using BAC probes (RP23-299L1 for *Dxz4* and RP23-338M16 for *Firre*) and an established protocol[14].

In situ DNase Hi-C data analysis

In situ DNase Hi-C was done on intact nuclei from wild-type and deleted/inverted Patski cells using a previously described method[45]. We sequenced the in situ DNase Hi-C libraries using paired-end reads 80bp and 150bp in length (Supplementary Table S2). For each *in situ* DNase Hi-C library, we mapped each end of the paired-end reads separately to the BL6 genome using the NCBI build v38/mm10 reference genome assembly obtained from the UCSC Genome Browser[46] and a pseudo-*spretus* genome using BWA-MEM (v0.7.3) in single-end mode using default parameters[47,48]. A pseudo-*spretus* genome was assembled by substituting available SNPs (from Sanger Institute, SNP database 2014/10/27 v4) into the BL6 reference genome, as described[34]. We retained only primary reads with MAPQ ≥30 for further analysis. Using heterozygous SNPs between the BL6 genome and the pseudo-*spretus* genome that were validated for our particular Patski cell line, we segregated all high-quality uniquely mapped reads (MAPQ ≥30) using an approach very similar to that described in[6], with the exception that in order to maximize the number of reads assigned to either the BL6 Xi or the *spretus* Xa, we required that only one end of the read be specifically mapped to one mouse species, while the

other end was allowed to be ambiguously mapped. This approach, commonly used to analyze data from hybrid systems, is based on the assumption that intrachromosomal contacts are much more frequent than interchromosomal contacts[8,10]. Briefly, each end of each read pair was assigned to one of three categories: (1) BL6-SNP reads containing only BL6-specific SNP(s); (2) *spretus*-SNP reads containing only *spretus*-specific SNP(s); (3) ambiguous reads that did not contain valid SNPs or that contained valid SNPs from both alleles. We refer to both BL6-SNP reads and *spretus*-SNP reads as "allele-specific reads", and reads that do not contain valid SNPs as "allele-uncertain reads". Reads were paired with their corresponding mates, and those read pairs with at least one end being allele-specific were retained for subsequent allele-specific analysis. To eliminate the bias due to the PCR duplication step, we removed redundant paired-end reads defined as those pairs where both ends were mapped to identical locations in the same genome assembly. This resulted in a set of valid read pairs representing DNA-DNA interactions (Supplementary Table S2).

We used the resulting valid read pairs to generate allele-specific whole-genome contact maps at 500kb and 40kb resolution. To do so, we partitioned the genome into non-overlapping bins and counted the number of allele-specific contacts (uniquely mapped valid paired-end reads) observed between each pair of bins. The dimension of the resulting allelic contact maps is the total number of bins in the genome, where entry (i, j) contains the contact count between bins i and j. We normalized the allele-specific contact maps using an iterative correction method described by[49] to obtain a filtered contact map with near-equal row and column sums. Prior to applying the iterative correction procedure, the contact maps were filtered as follows: bins along the diagonal, super-diagonal (+1 off-diagonal) and sub-diagonal (−1 off-diagonal) (representing entries dominated by self-ligation products) and bins with the lowest 2 % read coverage (representing sparsely populated regions dominated by spurious contacts) had their contact counts set to zero. Additionally, in order to better compare contact maps across multiple samples, intrachromosomal contact maps were quantile normalized to one another.

Inference of the 3D structure of X chromosomes

We inferred the 3D structure of the Xa and Xi chromosomes, separately, using the Pastis software[50] as described previously by[6], except that Pastis was run in PM2 mode rather than PM1. Briefly, each X chromosome is modeled as a series of n beads on a string, where n equals the number of bins in the X chromosome contact map. The model used by Pastis assumes that the observed contact count $C_{i,j}$ between beads i and j follows a Poisson distribution. The Poisson parameter of $C_{i,j}$ is a decreasing function of $d_{ij}(X)$ of the form $\beta d_{ij}(X)\alpha$, and $d_{ij}(X) = ||x_i - x_j||$ is the Euclidean distance between the beads i and j. Both α and β are inferred in the resulting optimization problem.

Coverage score analysis

To count the number of reads spanning each locus along the chromosome across all length scales we made a minor adaptation to the coverage score developed by[31]. Briefly, the coverage score for each bin at a particular resolution was calculated as the average number of interactions within bins spanning this central bin of interest. This score calculation can be visualized as sliding a V-shaped region (with the arms of the V extending all the way to the edges of the contact map) along the diagonal of the contact map and calculating the mean interaction counts (sum of counts/number of bins) within the region. One point of difference compared to Eser at al.[31] was that instead of normalizing the coverage score for each chromosome to fall within the range [0; 1], we normalized the scores by calculating log2(coverage score/chromosomal mean), as is done in the insulation score calculation (see below). Additionally, we did not restrict our analysis to interactions less than 100kb, but instead considered all interactions spanning each locus (i.e. the arms of the V-shaped region extend all the way to the borders of the contact map). However, when calculating the coverage scores for the unmerged data sets (WT, Del-hinge, Del-Dxz4, Del-Ds-TR, and Inv-5'Ds-TR), we did exclude bins within the first and last 10Mb of the chromosome in order to avoid extreme edge effects that were evident especially in the sparser samples. Furthermore, a five-window, degree two polynomial Savitzky-Golay filter was applied to the resulting vector to smooth the signal, with interpolated values being assigned to edge-case bins. A standardized smoothed coverage score (coverage z-score) or a coverage score rescaled to lie within

the [0; 1] interval were used where described in text.

Insulation score and domain boundary analysis

To determine regions of more localized interaction (domains) and associated boundaries, we implemented our own version of the insulation score method introduced by[32]. Briefly, at a particular resolution the insulation score for each bin along the chromosome was calculated as the average number of interactions falling within a square centered around each bin. This score calculation can be visualized as sliding a square along the diagonal of the contact map and calculating the mean interaction counts (sum of counts/number of bins) within the square. Bins at each end of the contact map, around which a complete square could not be formed, were not assigned an insulation score. A 3.5-by-3.5Mb square (7x7 bins) was used at 500kb resolution and a 520-by-520kb (13x13 bins) square was used at 40kb resolution. A five-window, degree two polynomial Savitzky-Golay filter was applied to the resulting vector to smooth the signal, with edge-case bins receiving interpolated values. A standardized smoothed insulation score (insulation z-score) was used where described in the text.

Domain boundaries were determined from the standardized insulation score (z-score) as local minima that showed at least a 1% change in the insulation z-score relative to both neighboring local maxima. I.e. Let Y be the local minima, X and Z be the neighboring maxima, and R = max insulation z-score – min insulation z-score. Then the local minima, Y, was designated as a domain boundary, if Y-X > 0.01 * R and Z-X > 0.01 * R.

CTCF ChIP-seq analysis

ChIP-seq was performed on WT and Del-hinge Patski cells using an antibody for CTCF (Upstate/Millipore) and an established protocol[51]. CTCF ChIP and input libraries were sequenced as paired-end reads of 75bp in length (Supplementary Table S4). CTCF and input ChIP-seq paired-end reads were mapped to the BL6 genome using the NCBI build v38/mm10 reference genome assembly obtained from the UCSC Genome Browser[46] using BWA-MEM (v0.7.3) in paired-end mode using

default parameters[47,48]. Primary mapped valid paired reads with MAPQ>=10 were de-duplicated and used for calling CTCF peaks based on biallelic reads using MACS2[52] (Supplementary Table S4). CTCF peaks were called using all reads (unsegregated) after normalization to their chromatin inputs (Supplementary Table S4). Over 80% of these peaks contained a validated BL6/*spretus* SNP (Supplementary Fig. S5b,c), with the vast majority covered by at least 5 reads and considered for allelic analysis (Supplementary Fig. S5d-g; see Methods). CTCF peaks containing *spretus*/BL6 SNPs that were covered by a total of at least 5 reads were considered for allelic analysis (Supplementary Fig. S5; Supplementary Table S4). This level of coverage was chosen because the overall allelic proportion (*spretus*/(*spretus*+BL6)) did not change much regardless of whether 5x or 10x coverage thresholds were used (Supplementary Fig. S5g). For each CTCF peak with sufficient SNP coverage, an allelic proportion of SNP read coverage (*spretus*/(*spretus*+BL6)) was calculated. Note that the small difference between autosomal allelic proportion distributions is due to aneuploidy for chromosomes 3 and 4 in some Patski cell clones, and disappears when these chromosomes are excluded from the analysis (Fig. 5a; Supplementary Fig. S5h). Using autosomal distributions of allelic proportion as a guide, peaks with an allelic proportion of greater than 0.7 were designated as *spretus*/Xa-specific, while those with an allelic proportion of less than 0.3 were deemed to be BL6/Xi-specific. Peaks with an allelic proportion falling within the range [0.3; 0.7] were considered biallelic (Supplementary Fig. S5; Supplementary Table S4). Subtracting one from allelic proportions results in 'd-scores' as previously described[8,53], with values ranging from -0.5 to +0.5. Peaks with positive d-scores are covered by more reads emanating from the *spretus*/Xa allele, while those peaks having negative d-scores show a BL6/Xi bias.

ATAC-seq analysis

ATAC-seq was done on WT and Del-hinge Patski cells using a published method[54]. ATAC-seq libraries were sequenced in multiple runs as paired-end reads of 75 or 150bp in length (Supplementary Table S5). ATAC-seq read mapping, peak calling, allelic assignment and d-score calculation were performed

as described for CTCF peaks, except that ATAC peaks were called without using an input (background) library (Supplementary Fig. S7; Supplementary Table S5). Fewer ATAC peaks contained SNPs than CTCF peaks (Supplementary Fig. S8a,b), which could be due to the fact that in both WT and Del-hinge many more, smaller ATAC peaks were called compared to CTCF peaks, possibly a consequence of not having background data to compare against as one does for ChIP-seq with input samples. The overall allelic proportion (*spretus*/(*spretus*+BL6)) for ATAC peaks did not change much regardless of whether 5x or 10x coverage was required (Supplementary Fig. S8e,f).

RNA-seq analysis

RNA-seq was done on WT, Del-hinge (clone a and b), Del-Dxz4, and Del-Ds-TR cells, as well as on Inv-Dxz4 and the WT subclone Patski2-4 (used to derive Inv-Dxz4) as described[14]. Induce DNA demethylation a 48h treatment using 4μM 5-aza-2'-deoxycytidine (5-aza-2dC) was used, followed by a 24h recovery period, reported to lead to ~2% of cells with reactivation of an X-linked GFP reporter gene[35]. We found that 5-aza-2dC treatment caused ~20% more cell death in WT and Del-hinge cells compared to the mock treatment, similar to that reported for dermal fibroblasts[35]. RNA-seq libraries were sequenced as paired-end reads of 75bp in length (Supplementary Table S6). RNA-seq paired-end reads were mapped to the UCSC mm10 (NCBI build v38) refSeq transcriptome[46] as downloaded and packaged in the iGenomes reference sequences and annotation files on July 17, 2015. (https://support.illumina.com/sequencing/sequencing_software/igenome.html). Tophat2 (v 2.0.12) (calling bowtie2 (v2.2.3)) was used to perform paired-end mapping allowing 6 mismatches but otherwise default parameters[55,56]. To determine biallelic expression levels, mapped reads were assigned to refSeq genes using HT-seq[57] and counts were converted into TPMs using custom R scripts.

Genes containing *spretus*/BL6 SNPs that were covered by a total of at least five reads were considered for allelic analysis. For each gene with sufficient SNP coverage, an allelic proportion of SNP read coverage (*spretus*/(*spretus*+BL6)) was calculated. Read counts for each gene were then

distributed to each allele based on this SNP-read allelic proportion, allowing us to perform differential expression between samples for each allele. Differential expression analysis was performed using DESeq2[58]. For the WT versus Del-hinge comparison, WT and Del-hinge control samples from the 5-aza-2dC experiment were pooled and treated as biological replicates for the WT and Del-hinge clone a samples.

Genes were deemed to escape XCI in the Patski WT line if their expression levels met the similar criteria to those used by Berletch et al.[34] in 2/3 of the WT samples (WT and the two WT 5-aza-2dC untreated replicates) (Supplementary Table S7): (1) the 99% lower confidence limit (alpha = 0.01) of the escape probability was greater than 0.01 based on a binomial distribution parameterized by the expected proportion of reads from the Xi indicating some contribution from the Xi; (2) the diploid gene expression measured by TPM >=1, indicating that the gene was expressed; (3) the Xi-TPM was >= 0.1, representing sufficient reads from the Xi, and (4) SNP coverage >= 5.

Acknowledgements. This study was supported by grants U54DK107979 (JS and WSN) and GM046883 (CMD). We thank C. Lee for DNA sequencing.

Figure legends

Fig. 1: *Dxz4* alone is sufficient to maintain the bipartite structure of the Xi. **a**. Contact maps are shown at 500kb resolution for the Xi in WT, Del-hinge (Xi deletion nt75637519-75764753), Del-Dxz4 (Xi deletion nt75721096-75764754), Inv-Dxz4 (Xi inversion nt75721096-75764754), Del-Ds-TR (Xi deletion nt75637501-75674037), and Inv-5' Ds-TR (Xi inversion nt75674046-75674952). The position of *Dxz4* on the mouse X chromosome and schematics of the allele-specific deletions/inversions are shown under the maps. See Supplementary Fig. S2a for contact maps of the Xa in the same clones. **b**. Relative position of the loci within the hinge region and location of the CRISPR/Cas9 induced

alterations. Arrows indicate the orientation of CTCF motifs at *Dxz4*[11] and at the 5'end of *Ds-TR*.

Fig. 2: *Dxz4* deletion or inversion changes contact distribution on the Xi. **a**-**c**. Contact maps (500kb resolution) for the Xi in WT* (a), Del-hinge/Dxz4 (b) and Inv-Dxz4 (c) visualized using Pearson correlation to highlight contact probabilities. **d**-**f**. 3D models based on the contact maps, show loss of the hinge in the Del-hinge/Dxz4 Xi, but not in the Inv-Dxz4 Xi in which *Dxz4* is pulled into the centromeric domain, whereas it is normally pulled into the telomeric domain in WT*. The position of *Firre* is indicated. **g**-**i**. Contact maps (500kb resolution) for 50Mb around the *Dxz4* locus to highlight the loss of superdomain structure in Del-hinge/Dxz4 and the shift in contacts in Inv-Dxz4, where a new contact domain forms between *Firre* and *Dxz4*. **j.** Allelic principal component (PC) score profiles for the Xa and Xi chromosomes in WT*, Del-hinge/Dxz4, and Inv-Dxz4, based on distance-corrected, normalized contact maps with counts binned at 500kb resolution. For the Xa, the top three allelic PC scores are shown (red, blue and green, respectively), matching the color scheme for autosomes in Supplementary Fig. S3. For the Xi, PC1 score profiles are plotted in black to emphasize that PC1 captures the bipartite structure rather than A/B compartments in WT*. PC2-4 are shown in red, blue and green, respectively. **k**. The variance explained by the top five allelic principal components for the Xa and Xi for WT*, Del-hinge/Dxz4, and Inv-Dxz4. **l.** Pairwise spearman correlation values and associated scatterplots between allelic PC1 scores for the Xa and Xi in WT*, Del-hinge/Dxz4, and Inv-Dxz4.

Fig.3: Unidirectional disruption of contacts on the Xi after *Dxz4* deletion or inversion. **a**. Differential contact map based on Pearson correlation transformed data at 500kb resolution to highlight differences between Del-hinge/Dxz4 Xi and WT* Xi (loss or gain of contacts in the Del-hinge/Dxz4 versus WT* appear blue or red, respectively). See also Supplementary Fig. S3d,e for comparison with differential contact maps based on untransformed count data. **b**. As in (a) to highlight differences between Inv-Dxz4 Xi and WT* Xi. See also Supplementary Fig. S3f,g for comparison with differential contact maps

based on untransformed count data. **c**. Virtual 4C plots derived from Hi-C data at 500kb resolution using a 500kb region around *Dxz4* as the viewpoint on the Xi in WT* (blue), Del-hinge/Dxz4 (red), and Inv-Dxz4 (black). Y-axis (contact counts) limited to 20% of maximum. The positions of *Firre*, *Dxz4* and *Xist* are indicated. **d**. Standardized coverage score profiles at 500kb resolution for the Xi in WT* (blue), Del-hinge/Dxz4 (red), and Inv-Dxz4 (black). The positions of *Firre*, *Dxz4* and *Xist* are indicated. **e.** Coverage scores (rescaled to interval [0; 1]) at 40kb resolution within a 8Mb region around the midpoint of *Dxz4* for Xi in WT* (blue), Del-hinge/Dxz4 (red), and Inv-Dxz4 (black). The horizontal red dashed line indicates the threshold (0.015) used to define the Xi superdomain boundary in WT* and Inv-Dxz4 representing a region across which minimal interactions occur. The transparent blue background highlights the Xi boundary region in WT* and the transparent gray background highlights the right-shifted Xi boundary in Inv-Dxz4.

Fig. 4: Changes in TAD configuration on the Xi after *Dxz4* deletion or inversion. **a**. Comparisons of contacts on the Xa and Xi in WT*, and Del-hinge/Dxz4 at individual loci. Changes in TAD configuration at 40kb resolution are shown within 4Mb regions, each centered at a specific gene, including two genes normally subject to XCI, *Edar2* and *Zfx*, and a gene that escapes XCI, *Ddx3x*. **b**. Standardized insulation Z-score profiles at 500kb resolution for the whole Xi in WT* (blue), Del-hinge/Dxz4 (red), and Inv-Dxz4 (black). Insulation scores were derived based on the average number of contacts within a 3.5Mb sliding window (seven 500kb bins), considering a 10Mb swath along the diagonal of the contact maps. The positions of *Firre*, *Dxz4*, *Xist*, *Eda2r* and *Zfx* are indicated. **c**. Genome browser plots showing insulation Z-scores and TADs along a 10Mb region of the Xi (red) and Xa (blue), centered around *Dxz4* (arrow) in WT*, Del-hinge/Dxz4, and Inv-Dxz4. **d**. Violin plots showing the distribution of standardized insulation Z-scores at each 500kb bin centered within a 3.5Mb sliding window along the Xi and Xa in WT (blue), Del-hinge (red), Del-Dxz4 (orange), Inv-Dxz4 (black), Del-Ds-TR (green), and Inv-5'Ds-TR cells (indigo). **e**. Hierarchical clustering based on the Euclidean distance between standardized insulation Z-scores using 500kb bins along the Xi and Xa in WT, Del-hinge, Del-Dxz4, Inv-

Dxz4, Del-Ds-TR, and Inv-5'Ds-TR cells.

Fig. 5. Xi CTCF and ATAC peak distributions change in Del-hinge cells. **a**. Density histograms of the distribution of allelic proportions of CTCF peaks (*spretus*/(*spretus* +BL6) along autosomes and the X-chromosomes for WT (blue) and Del-hinge (red). The modes of the X-chromosome distributions are given in parentheses. *Chromosomes 3 and 4 were removed from the autosomes because they show aneuploidy (see also Supplementary Fig. S5). **b**. Percentages of CTCF peaks in WT (blue) and in Del-hinge (red) along the autosomes and the X-chromosomes classified as *spretus*-specific, BL6-specific, or biallelic peaks. **c**. Genome browser tracks of CTCF peaks on the Xa and Xi, of CTCF peak d-scores ((Xa/(Xa+Xi) – 0.5), and of CTCF peaks assigned as Xa-specific, biallelic, or Xi-specific for WT (blue) and Del-hinge (red) along a region of the X-chromosome that includes *Zfx* (a gene subject to XCI in WT, which reactivated in Del-hinge) and *Eif2s3x* (a gene that escapes XCI). CTCF peaks that appear on the Xi in a region around *Zfx* in Del-hinge are indicated with arrows. **e**-**g**. Same analyses as in **a**-**c** for ATAC peaks.

Fig. 6. No change in Xi location relative to the nuclear periphery or the lamina after *Dxz4* deletion or inversion. **a**. Examples of nuclei stained with DAPI (blue) and immunostained using an H3K27me3 antibody to locate the Xi (red), and an antibody to nucleophosmin to locate the nucleolus (green). The Xi was either located at the periphery (i), near the nucleolus (ii), or sandwiched between the periphery and the nucleolus (iii) (scored in both categories in **b**), or close neither to the periphery nor to the nucleolus (iv). **b**. Percentage of nuclei with the Xi near the periphery, the nucleolus, and neither of these locations in WT, Del-hinge and Inv-Dxz4.

Fig. 7. Gene expression changes after deletion of the hinge. **a**. Plots of expression fold-changes (log2) between Del-hinge versus WT for transcripts on the X chromosome and for subsets of transcripts

specifically from the Xi and Xa, based on SNPs, relative to expression levels (mean read counts). The total number of genes examined and the number of genes that show increased expression (Up) or lower expression (Down) in Del-hinge versus WT is indicated. **b**. Box plots showing the distribution of fold changes in allelic gene expression in Del-hinge versus WT for autosomal (A_spretus or A_BL6) and X-linked (Xa or Xi) genes that had increased expression in WT or in Del-hinge. Red, *spretus* alleles; blue, BL6 alleles. **c**. Distribution of genes with expression fold changes between Del-hinge versus WT along the length of the Xi, considering 209 genes normally subject to XCI (orange) and 29 genes that escape XCI (green). Only genes with >5 mean read counts from WT and Del-hinge were included. The position of *Dxz4* is indicated. **d**. Models for the role of *Dxz4* in establishing long-range contacts with other loci. Diagram of the *Dxz4* locus (black) with adjacent centromeric (red) and telomeric (green) superdomains of the Xi. The orientation of CTCF binding motifs is shown with black arrows, with 14 motifs being represented. Potential CTCF sites located in the telomeric superdomain are shown as green arrows, and on the centromeric domain as red arrows. In WT cells contacts between *Dzx4* and loci telomeric to the locus would result in the formation of loops anchored at *Dxz4* by the correct alignment of CTCF motifs, which would stall cohesin rings that continuously extrude loops (not depicted). *Dxz4* would be pulled to the telomeric end of the hinge. After *Dxz4* inversion contacts would shift to the centromeric superdomain, and be especially enhanced between *Dxz4* and *Firre*. A new de-condensed hinge would form, with *Dxz4* located at its centromeric end. Note that in WT and Inv-Dxz4, loops are depicted as anchored at each CTCF binding site on *Dxz4*. However, the process of loop formation is not static, but rather highly dynamic; thus at a given time some of the loops would not be anchored and larger or smaller loops may form. In addition, one CTCF molecule at *Dxz4* may be sufficient to stall cohesin, instead of two as depicted here.

Supplementary Figure legends

Supplementary Fig. S1. Generation and verification of the deletions and inversions of the mouse Xi. **a**.

Schematic of the hinge region indicating the position of the guide RNAs used for CRISPR/Cas9 editing. Deletion of the whole hinge (127kb nt75637519-75764753) was obtained in two independent experiments using Ds1 and Dx2 and includes the transcriptional start site of the *Dxz4*-associated lncRNA gene *4933407K13Rik*; Deletion of *Ds-TR* (37kb nt75637501-75674037) was obtained using Ds1 and Ds2 and does not include the promoter region of *Ds-TR*; Inversion (907bp nt75674046-75674952) of 2 of 3 CTCF binding sites located 5' of *Ds-TR* was obtained using Ds2 and Ds3; Deletion and inversion of *Dxz4* (44kb nt75721096-75764754) were obtained using Dx1 and Dx2 (Supplementary Table S1). **b**. Example of a verification of one of the alterations: PCR amplification using primers F1 and R1 followed by Sanger sequencing verified loss of SNPs from the BL6 allele and PCR amplification using primers F1 and R2 revealed the new junction sequence (see also Supplementary Table S2). **c**. Deletion of the hinge was also verified by fluorescence in situ hybridization (FISH) using a BAC probe for *Dxz4* (red) and a control BAC probe for *Firre* (green). Left, example of a metaphase chromosome preparation with one intact X chromosome with green and red signals (arrows) and a *Dxz4*-deleted X chromosome with only a green signal (arrow); Right, example of a nucleus with one red and two green signals (arrows).

Supplementary Fig. S2. **a**. Contact maps for the Xa do not differ between WT and Del-hinge, Del-Dxz4, Inv-Dxz4, Del-Ds-TR, and Inv-5'Ds-TR. Contact maps are shown at 500kb resolution. The position of *Dxz4* on the X and schematics of the allele-specific deletions/inversion are shown under the maps (see also Fig. 1a). **b.** Contact maps for the Xi and Xa representing two pooled sets of Hi-C contacts: a data set representing wild-type and Del-Ds-TR (designated WT*), and another representing Del-hinge and Del-Dxz4 (designated Del-hinge/Dxz4). Allelic contact maps for each pool are very similar to those obtained from the pooled data (see Fig. 1a for comparison to maps generated for each deletion). Contact maps were generated at 500kb resolution. **c**. 3D models of the Xa in WT* and Del-hinge/Dxz4 (see Fig. 2d-f for the 3D models of the Xi).

Supplementary Fig. S3. Allelic principal component (PC) score profiles for autosomes based on distance-corrected, normalized contact maps. **a.** WT*, Del-hinge/Dxz4, and Inv-Dxz4 allelic principal component (PC) score profiles for an exemplar autosome (chromosome 2) based on distance-corrected, normalized contact maps with counts binned at 500kb resolution (*spretus* top row; B6 bottom row). In each case, the top three allelic PC scores are shown (red, blue and green, respectively). **b.** The variance explained by the top five allelic principal components averaged across all autosomes chromosomes (*spretus* top row; B6 bottom row) for WT*, Del-hinge/Dxz4, and Inv-Dxz4. **c.** Pairwise spearman correlation values and associated scatterplots between allelic PC1 scores for autosomes concatenated end-to-end for WT*, Del-hinge/Dxz4, and Inv-Dxz4. See Fig. 2j-l for similar analyses on the X chromosomes. **d**. Differential contact map based on untransformed count data at 500kb resolution to highlight differences between Del-hinge/Dxz4 Xi and WT* Xi (loss or gain of contacts in the Del-hinge/Dxz4 versus WT* appear blue or red, respectively). **e**. As in (d) for differential contact map based on Pearson correlation transformed data. **f**. As in (d) to highlight differences between Inv-Dxz4 Xi and WT* Xi. **g**. As in (f) for differential contact map based on Pearson correlation transformed data. **h**. Virtual 4C plots derived from the Hi-C data for various 500kb viewpoints positioned along the Xi for WT* (blue), Del-hinge/Dxz4 (red), and Inv-Dxz4 (black). Y-axis (contact counts) limited to 20% of maximum. The position of the viewpoints is indicated. See text for details and additional data in Fig. 3.

Supplementary Fig. S4. Coverage-score analyses of all cell lines. **a**, **b**. Standardized coverage Z-score profiles using 500kb resolution Hi-C contact map data along the Xi (a) and Xa (b) in WT (blue), Del-hinge (red), Del-Dxz4 (orange), Inv-Dxz4 (black), Del-Ds-TR (green), and Inv-5'Ds-TR cells (indigo). **c**. Violin plots showing the distribution of standardized coverage Z-scores using 500kb bins for the Xi and Xa in WT (blue), Del-hinge (red), Del-Dxz4 (orange), Inv-Dxz4 (black), Del-Ds-TR (green), and Inv-5'Ds-TR cells (indigo). **d**. Hierarchical clustering based on the Euclidean distance between standardized coverage Z-scores using 500kb bins along the Xi and Xa in WT, Del-hinge, Del-Dxz4, Inv-Dxz4, Del-Ds-TR, and Inv-5'Ds-TR cells. **e**. Hierarchical clustering based on Pearson correlation (using

1 – *r* as the distance measure) for standardized coverage Z-scores using 500kb bins along the Xi and Xa in WT, Del-hinge, Del-Dxz4, Inv-Dxz4, Del-Ds-TR, and Inv-5'Ds-TR cells. **f**. Profiles of insulation Z-scores (log2 insulation scores/mean then standardized to obtain a Z-score across the Xi in WT* (blue), Del-hinge/Dxz4 (red), and Inv-Dxz4 (black). Insulation scores were derived from 500kb bins as the average number of contacts within a 3.5Mb sliding window (seven 500kb bins). The positions of *Firre, Dxz4, Zfx, Eda2r*, and *Xist* are indicated. **g**. Profiles of insulation Z-scores as in (f) but using 40kb bins and a 520kb window (13 bins) for a region between the *Firre* and *Xist* loci for both Xi and Xa in WT* (blue) and Del-hinge/Dxz4 (red). **h-i**. Hierarchical clustering based on the Euclidean distance (h) and Pearson correlation (using 1 – *r* as the distance measure; i) between standardized insulation scores at 40kb resolution within a 520kb window considering the Xi and Xa in WT*, Del-hinge/Dxz4 and Inv-Dxz4. **j.** Hierarchical clustering based on the adjusted Rand index to quantify the correspondence between TADs called using insulation scores within a 520kb window at 40kb resolution along the Xi and Xa in WT*, Del-hinge/Dxz4 and Inv-Dxz4.

Supplementary Fig. S5. CTCF ChIP-seq SNP coverage and allelic ratios. **a**. PCR amplification of *Dxz4* and of the control autosomal gene *H19* was done on the input fraction (10% input), the CTCF ChIP fraction (CTCF ChIP) and the no antibody fraction (No Ab) for WT cells (+/+) and Del-hinge cells (Del-hinge/+). A strong decrease in CTCF enrichment at *Dxz4* is seen in Del-hinge cells due to deletion of *Dxz4* on the Xi, which normally binds CTCF. **b**. Histograms of SNP counts within CTCF peaks along autosomes and the X-chromosomes in WT (blue). **c**. As in (a), except for Del-hinge (red). **d**. Histograms of the SNP read coverage within SNP-containing CTCF peaks along autosomes and the X-chromosomes in WT (blue). **e**. As in (c), except for Del-hinge (red). **f**. CTCF peak counts at three different levels of read coverage (0, 5x, 10x) along autosomes and the X-chromosomes in WT (blue) and Del-hinge (red). **g**. As in (e), but showing allelic proportions. **h**. Density histograms of the distribution of allelic proportions of CTCF peaks (*spretus*/(*spretus* +BL6)) along all autosomes and the X-chromosomes for WT (blue) and Del-hinge (red). The modes of the X-chromosome distributions are

given in parentheses (see also Fig. 5a,b).

Supplementary Fig. S6. a. Genome browser tracks of CTCF and ATAC peak d-scores ((*spretus*/(*spretus* +BL6) – 0.5) and of peaks assigned as *spretus*-specific, common, or BL6-specific for WT (blue) and Del-hinge (red) peaks along chromosome X. **b**. Same analysis for two regions of the X chromosome: chrX: 65566000-82596000 and at the *Firre* locus.

Supplementary Fig. S7. a. Genome browser tracks of CTCF and ATAC peak d-scores ((*spretus*/(*spretus* +BL6) – 0.5) and of peaks assigned as *spretus*-specific, common, or BL6-specific for WT (blue) and Del-hinge (red) peaks along chromosomes 1, 10, and 19. **b**. Same analysis for two regions that include imprinted genes on chromosome 7.

Supplementary Fig. S8. ATAC-seq SNP coverage and allelic ratios. **a**. Histograms of SNP counts within ATAC peaks along autosomes and the X-chromosomes in WT (blue). **b**. As in (a), except for Del-hinge (red). **c**. Histograms of the SNP read coverage within SNP-containing ATAC peaks along autosomes and the X-chromosomes in WT (blue). **d**. As in (c), except for Del-hinge (red). **e**. ATAC peak counts at three different levels of read coverage (0, 5x, 10x) along autosomes and the X-chromosomes in WT (blue) and Del-hinge (red). **f**. As in (e), but showing allelic proportions. **g**. Density histograms of the distribution of allelic proportions of ATAC peaks (*spretus*/(*spretus* +BL6)) along all autosomes and the X-chromosomes for WT (blue) and Del-hinge (red). The modes of the X-chromosome distributions are given in parentheses (see also Fig. 5e,f).

Supplementary Fig. S9. Gene expression changes in Del-hinge clone a **(a)**, Del-hinge clone b **(b)**, Del-Dxz4 **(c)**, and Del-Ds-TR **(d)**, Inv-Dxz4 clone a compared to WT cells. Scatter plots of Xi- and Xa-specific expression between each deleted clone and WT are shown for genes with mean log2(TPM) >=

-4 diploid (i.e. mean TPM >= 0.0625). Dot lines represent 1.5-fold cutoffs.

Supplementary Fig. S10. Gene expression changes in Inv-Dxz4 (a), and in WT and Del-hinge after 5-aza-2dC treatment (b-e). **a**. Plots of expression fold-changes (log2) between Inv-Dxz4 versus WT (Patski2-4) for transcripts on the X chromosome and for subsets of transcripts specifically from the Xi and Xa, based on SNPs, relative to expression levels (mean read counts). The total number of genes examined and the number of genes that show increased expression (Up) or lower expression (Down) in Inv-Dxz4 versus WT is indicated. **b-c**. Plots of expression fold-changes (log2) between 5-aza-2dC -treated cells and controls in WT (b) and Del-hinge clone a (c) for all transcripts on the X chromosome (chrX) and for allele-specific transcripts from the Xi and Xa, relative to expression levels (mean read counts). The total number of genes examined and the number of genes that show increased expression (Up), decreased expression (Down) and the total number of genes with differential expression (DE) in treated versus untreated cells is indicated. **d**-**e**. Box plots showing the distribution of allelic expression of genes with increased expression (Up) in untreated and 5-aza-2dC treated cells in WT (d) and Del-hinge clone a (e) for autosomal (A_*spretus* or A_BL6) and X-linked (Xa or Xi) genes. Red, *spretus* alleles; blue, BL6 alleles.

Supplementary Table S1. List of sgRNAs for CRISP/Cas9 editing. *sgRNAs designed by CHOPCHOP were selected to include BL6 SNPs at the PAM site (red) if available. The *spretus* SNPs are listed as small letters in parenthesis. A pair of sgRNAs were used to edit each target: Ds-1&2 for Del-Ds-TR, Ds-2&3 for Inv-5' Ds-TR (CTCF peak inversion (chrX:75674066-75674317) at the 5' of Ds-TR, Dx1&2 for Del-Dxz4 and Inv-Dxz4, and Ds1&Dx2 for Del-hinge. See also Supplementary Fig. S1a.

Supplementary Table S2. In situ DNase Hi-C valid read pairs before and after merging. The number of valid read pairs and their length is listed for each Hi-C library together with the calculated total number

of reads for each individual cell type, and for the pooled data sets (WT* and Del-hinge/Dxz4). Reads are mapped to mm10 using BWA MEM. The number of BL6-specific and *spretus*-specific interactions based on one unambiguously mapped end and discarding interactions < 20kb are listed.

Supplementary Table S3. Summary of TADs. The estimated number of TADs is shown for the Xi and Xa for each cell line using different bin size.

Supplementary Table S4. Summary of CTCF ChIP-seq read counts and peaks. The number of read pairs is listed for the ChIP and the input in Patski WT and Del-hinge. The number of diploid CTCF peaks and SNP-containing CTCF peaks including genome-wide, autosomal and X-linked peaks are listed together with the percentage in each category and the ratio between Del-hinge and WT. The number of allelic peaks including covered peaks, Xa- and Xi-assigned peaks and common peaks are listed together with the percentages in each category and the ratio between Del-hinge and WT.

Supplementary Table S5. Summary of ATAC-seq read counts and peaks. As described in Supplementary Table S4, but for ATAC-seq.

Supplementary Table S6. RNA-seq read counts and alignment statistics. Total RNA-seq reads, mapped reads and their percentage of total are shown for each library.

Supplementary Table S7. Genes (mm10) called escape in 2/3 Patski WT lines. The genes listed were classified as genes that escape XCI in 3/3 WT lines based on criteria described in the text, except for *Ftx* classified as escape in 2/3 lines. For each line, the ratio between the number of reads from the Xi versus the total number of reads assigned either to the Xi or Xa is listed, together with the number of reads containing SNPs specific for the Xi or Xa, the total TPM, the Xi- and Xa- specific TPM, and the lower and upper confidence limits of escape probability (see methods).

Supplementary Table S8. Differential Xi-specific expression changes between WT and Del-hinge clone a. DESeq2 differential expression results between Del-hinge clone a and WT for each X-linked gene. Columns give the mean expression, log2 fold change in expression, standard error in log2 fold change, DE test statistic, p-value of test, and p-value adjusted to account for multiple testing. Highlighted genes show a significant differential expression (|log fold change| >= 0.5 and adjusted p-value <=0.05).

Supplementary Table S9. Differential Xi-specific expression changes between WT (Patski2-4) and Inv-Dxz4. As described in Supplementary Table S8, but for Inv-Dxz4 and Patski2-4 WT.

Supplementary Table S10. Differential Xi-specific expression changes between 5-aza-2dC untreated and treated WT. As described in Supplementary Table S8, but for 5-aza-2dC treated and untreated WT.

Supplementary Table S11. Differential Xi-specific expression changes between 5-aza-2dC untreated and treated Del-hinge clone a. As described in Supplementary Table S8, but for 5-aza-2dC treated and untreated Del-hinge clone a.

Supplementary Table S12. List of primers. *Five different combinations of primer pairs were used for each pair of cutting sites. For example, for Dxz4 cut1 and 2, Dx_F1/R1, F2/R2, F1/R2 (deletion), F1/F2 (inversion).

References

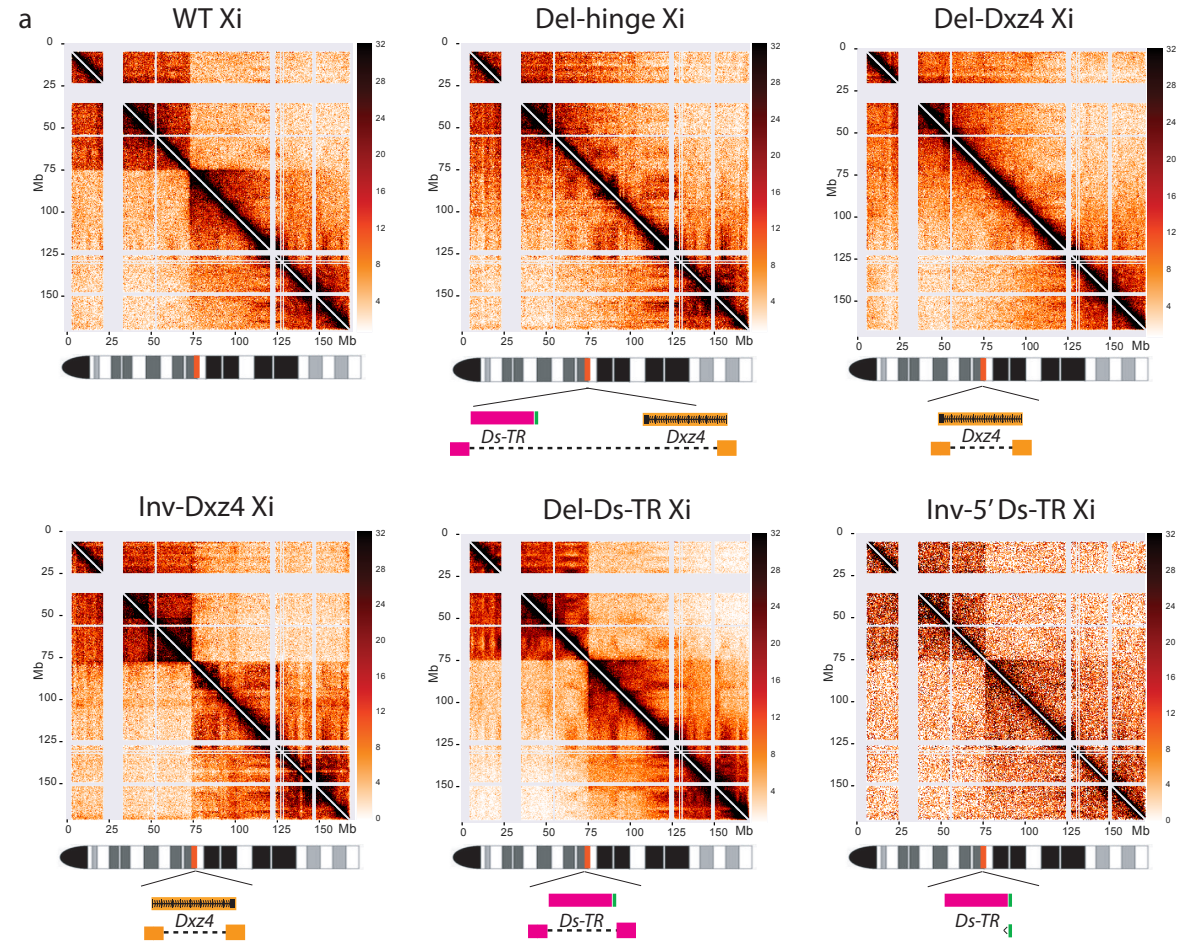
1. Heard, E. & Disteche, C.M. Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes Dev* **20**, 1848-67 (2006).

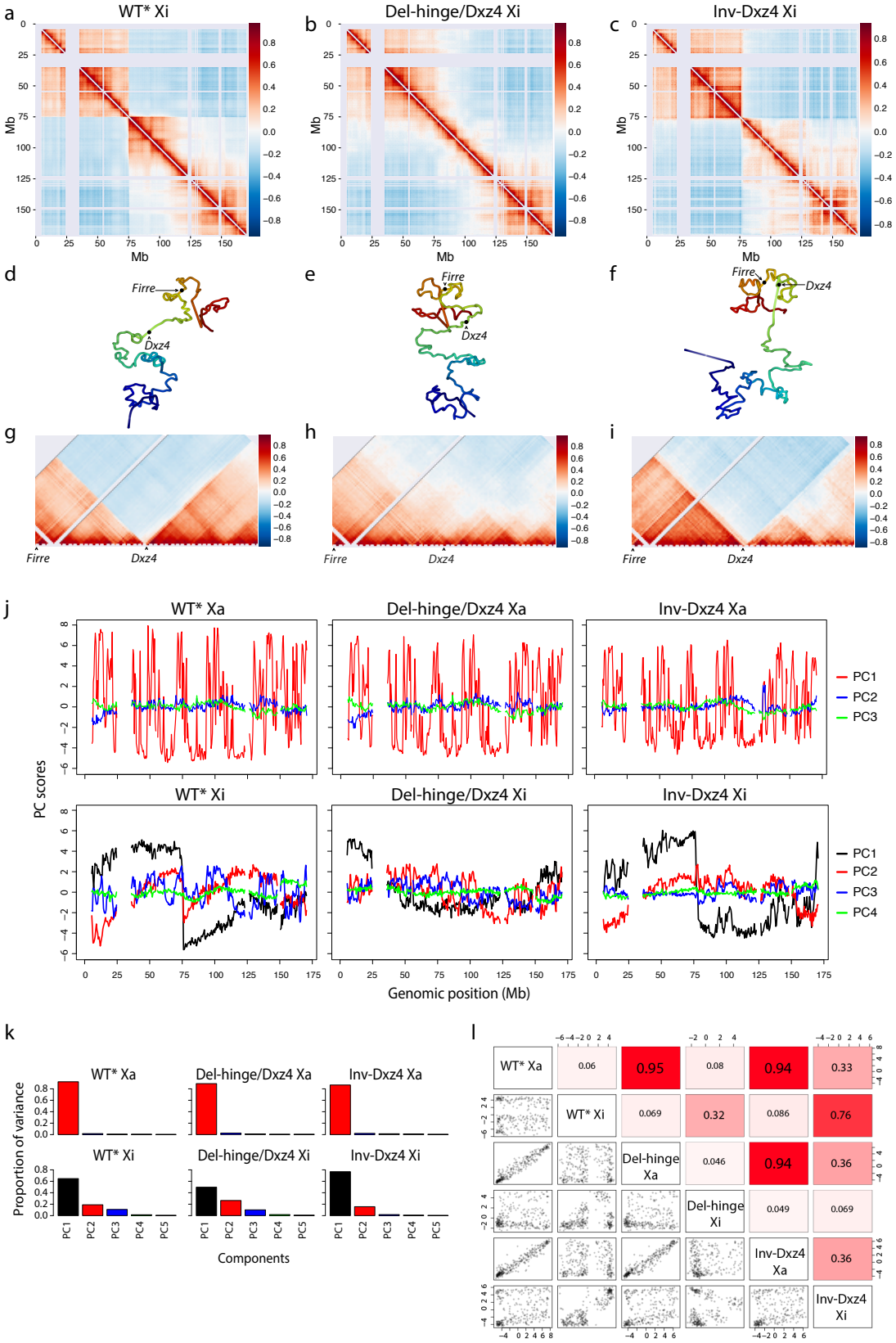
2. Vallot, C., Ouimette, J.F. & Rougeulle, C. Establishment of X chromosome inactivation and epigenomic features of the inactive X depend on cellular contexts. *Bioessays* **38**, 869-80 (2016).
3. Barr, M.L. & Bertram, E.G. A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. *Nature* **163**, 676 (1949).
4. Lyon, M.F. Sex chromatin and gene action in the mammalian X-chromosome. *Am J Hum Genet* **14**, 135-48 (1962).
5. Ohno, S. & Hauschka, T.S. Allocycly of the X-chromosome in tumors and normal tissues. *Cancer Res* **20**, 541-5 (1960).
6. Deng, X. et al. Bipartite structure of the inactive mouse X chromosome. *Genome Biol* **16**, 152 (2015).
7. Darrow, E.M. et al. Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proceedings of the National Academy of Sciences of the United States of America* **113**, E4504-12 (2016).
8. Giorgetti, L. et al. Structural organization of the inactive X chromosome in the mouse. *Nature* **535**, 575-9 (2016).
9. Minajigi, A. et al. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* **pii: aab2276**(2015).
10. Rao, S.S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-80 (2014).
11. Horakova, A.H. et al. The mouse DXZ4 homolog retains Ctcf binding and proximity to Pls3 despite substantial organizational differences compared to the primate macrosatellite. *Genome Biol* **13**, R70 (2012).
12. Horakova, A.H., Moseley, S.C., McLaughlin, C.R., Tremblay, D.C. & Chadwick, B.P. The macrosatellite DXZ4 mediates CTCF-dependent long-range intrachromosomal interactions on the human inactive X chromosome. *Hum Mol Genet* **21**, 4367-77 (2012b).
13. Chadwick, B.P. DXZ4 chromatin adopts an opposing conformation to that of the surrounding chromosome and acquires a novel inactive X-specific role involving CTCF and antisense transcripts. *Genome Res* **18**, 1259-69 (2008).
14. Yang, F. et al. The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. *Genome Biology*, 1-17 (2015).
15. Giacalone, J., Friedes, J. & Francke, U. A novel GC-rich human macrosatellite VNTR in Xq24 is differentially methylated on active and inactive X chromosomes. *Nat Genet* **1**, 137-43 (1992).
16. Ghirlando, R. & Felsenfeld, G. CTCF: making the right connections. *Genes Dev* **30**, 881-91 (2016).
17. Dixon, J.R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-80 (2012).
18. Nora, E.P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-5 (2012).
19. Nora, E.P. et al. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944 e22 (2017).
20. Fudenberg, G. et al. Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep* **15**, 2038-49 (2016).
21. Sanborn, A.L. et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A* **112**, E6456-65 (2015).
22. Haarhuis, J.H.I. et al. The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell* **169**, 693-707 e14 (2017).
23. Tedeschi, A. et al. Wapl is an essential regulator of chromatin structure and chromosome segregation. *Nature* **501**, 564-8 (2013).
24. de Wit, E. et al. CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell* **60**, 676-84 (2015).
25. Guo, Y. et al. CRISPR Inversion of CTCF Sites Alters Genome Topology and

Enhancer/Promoter Function. *Cell* **162**, 900-10 (2015).
26. Ma, W. et al. Fine-scale chromatin interaction maps reveal the cis-regulatory landscape of human lincRNA genes. *Nat Methods* **12**, 71-8 (2015).
27. Lingenfelter, P.A. et al. Escape from X inactivation of Smcx is preceded by silencing during mouse development. *Nat Genet* **18**, 212-3 (1998).
28. Yang, F., Babak, T., Shendure, J. & Disteche, C.M. Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Res* **20**, 614-22 (2010).
29. Hacisuleyman, E. et al. Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat Struct Mol Biol* **21**, 198-206 (2014).
30. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-93 (2009).
31. Eser, U. et al. Form and function of topologically associating genomic domains in budding yeast. *Proc Natl Acad Sci U S A* **114**, E3061-E3070 (2017).
32. Crane, E. et al. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240-4 (2015).
33. Favorov, A. et al. Exploring massive, genome scale datasets with the GenometriCorr package. *PLoS Comput Biol* **8**, e1002529 (2012).
34. Berletch, J.B. et al. Escape from X inactivation varies in mouse tissues. *PLoS Genet* **11**, e1005079 (2015).
35. Vallender, T.W. & Lahn, B.T. Improved DNA methylation analysis via enrichment of demethylated cells expressing an X-inactivated transgene. *Biotechniques* **41**, 461-6 (2006).
36. da Rocha, S.T. & Heard, E. Novel players in X inactivation: insights into Xist-mediated gene silencing and chromosome conformation. *Nat Struct Mol Biol* **24**, 197-204 (2017).
37. Almeida, M. et al. PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation. *Science* **356**, 1081-1084 (2017).
38. Pohlers, M., Calabrese, J.M. & Magnuson, T. Small RNA expression from the human macrosatellite DXZ4. *G3 (Bethesda)* **4**, 1981-9 (2014).
39. Tang, Z. et al. CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell* **163**, 1611-27 (2015).
40. Filippova, G.N. et al. Boundaries between chromosomal domains of X inactivation and escape bind CTCF and lack CpG methylation during early development. *Dev Cell* **8**, 31-42 (2005).
41. Chadwick, B.P. & Willard, H.F. Chromatin of the Barr body: histone and non-histone proteins associated with or excluded from the inactive X chromosome. *Hum Mol Genet* **12**, 2167-78 (2003).
42. Hooper, M., Hardy, K., Handyside, A., Hunter, S. & Monk, M. HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature* **326**, 292-295 (1987).
43. Labun, K., Montague, T.G., Gagnon, J.A., Thyme, S.B. & Valen, E. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res* **44**, W272-6 (2016).
44. Montague, T.G., Cruz, J.M., Gagnon, J.A., Church, G.M. & Valen, E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res* **42**, W401-7 (2014).
45. Ramani, V. et al. Mapping 3D genome architecture through in situ DNase Hi-C. *Nat Protoc* **11**, 2104-21 (2016).
46. Kent, W.J. et al. The human genome browser at UCSC. *Genome Res* **12**, 996-1006 (2002).
47. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-95 (2010).
48. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*, 1303.3997 (2013).
49. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat Methods* **9**, 999-1003 (2012).
50. Varoquaux, N., Ay, F., Noble, W.S. & Vert, J.P. A statistical approach for inferring the 3D structure of the genome. *Bioinformatics* **30**, i26-33 (2014).

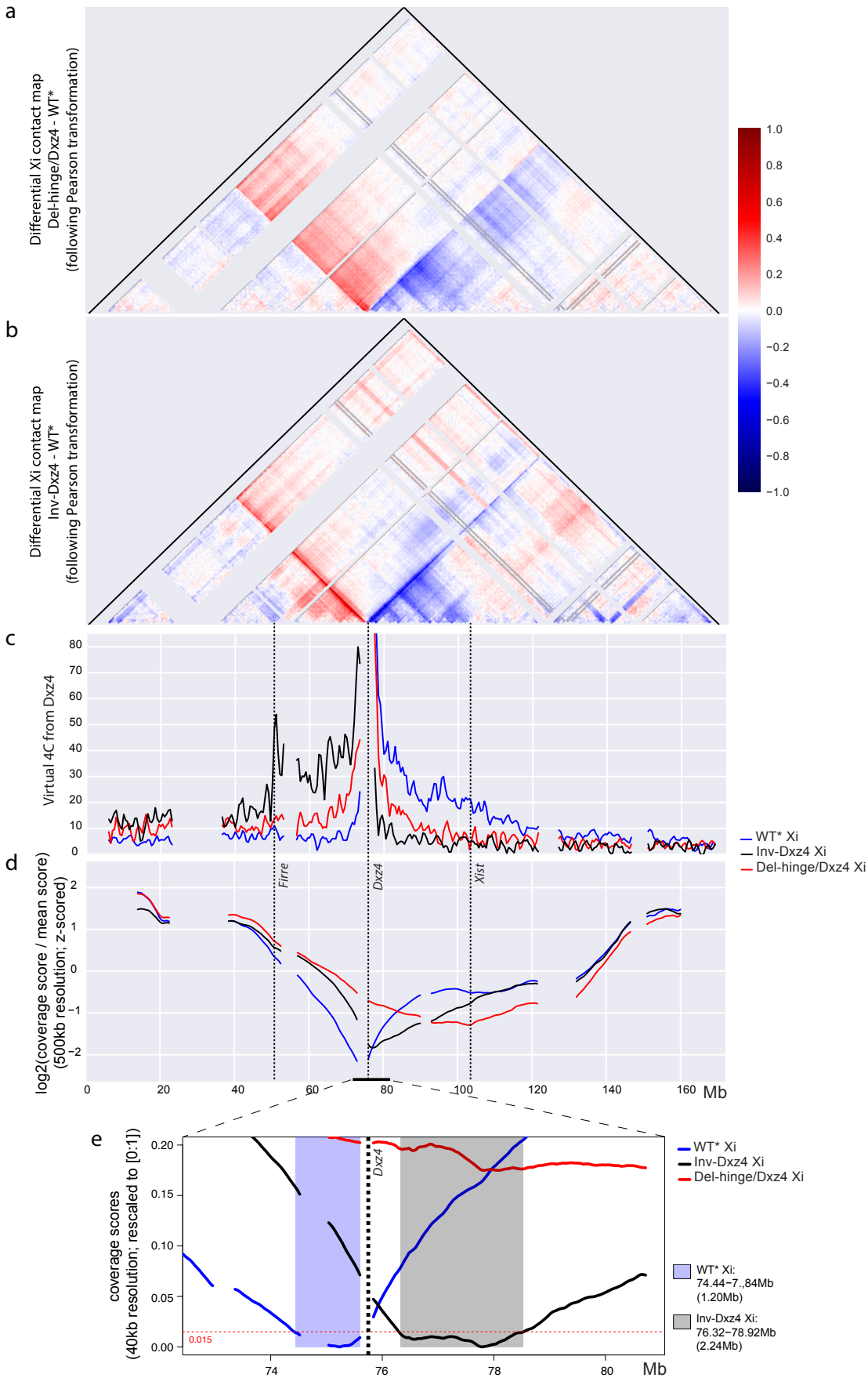
51. Deng, X. et al. Mammalian X upregulation is associated with enhanced transcription initiation, RNA half-life, and MOF-mediated H4K16 acetylation. *Dev Cell* **25**, 55-68 (2013).
52. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
53. Xu, J. et al. Landscape of monoallelic DNA accessibility in mouse embryonic stem cells and neural progenitor cells. *Nat Genet* **49**, 377-386 (2017).
54. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. & Greenleaf, W.J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-8 (2013).
55. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-11 (2009).
56. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36 (2013).
57. Anders, S., Pyl, P.T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-9 (2015).
58. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

Bonora Deng Figure 1

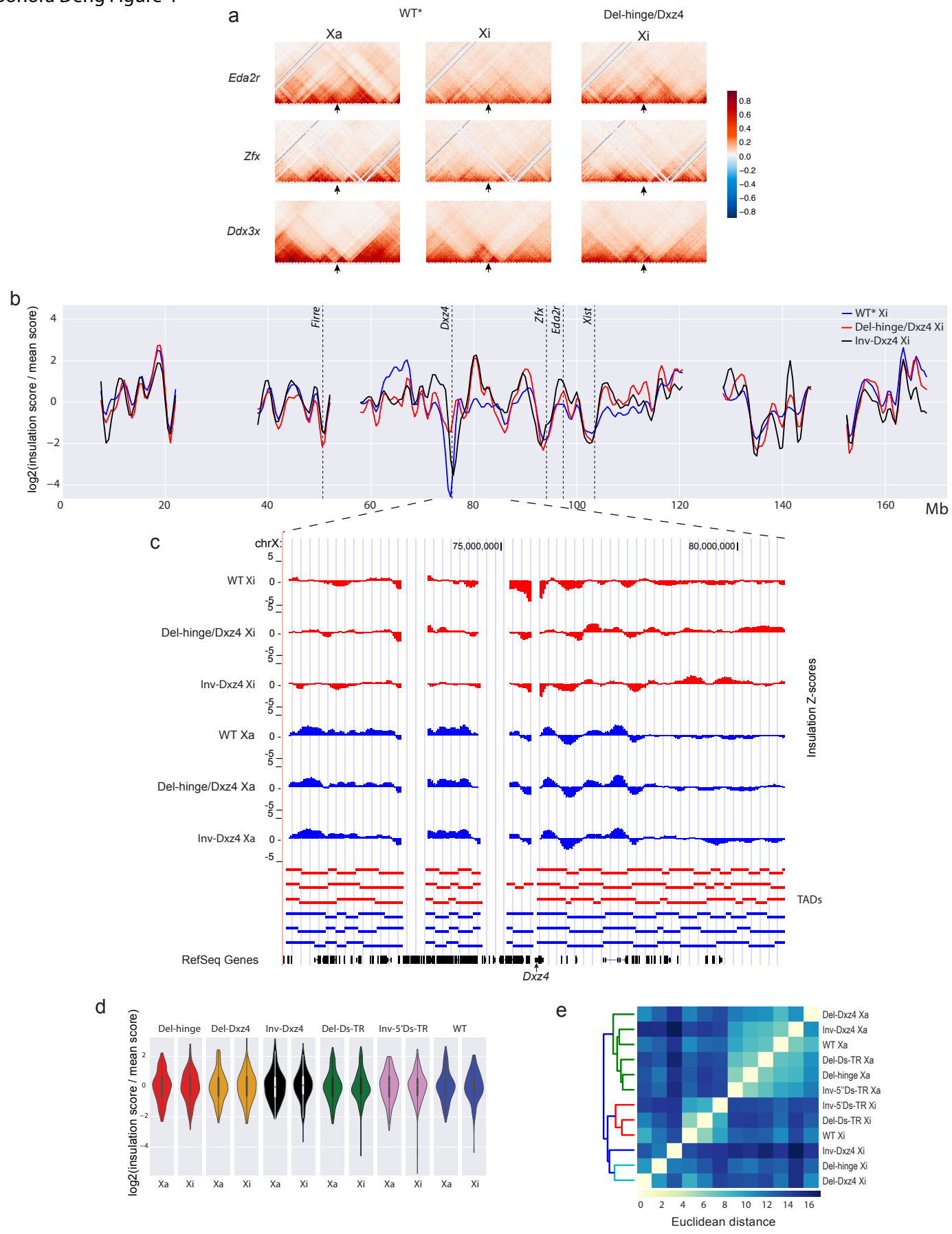




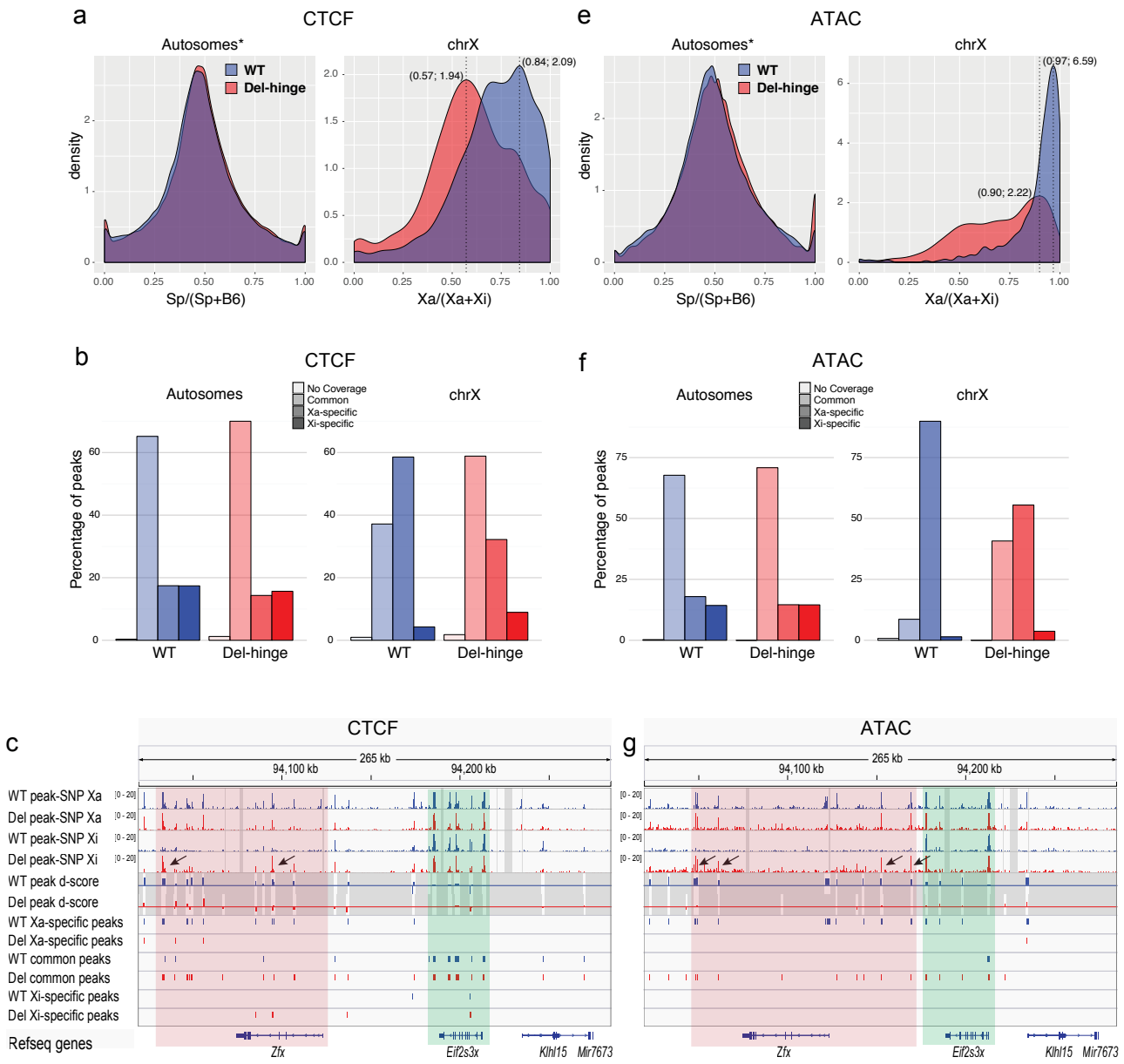
Bonora Deng Figure 3



Bonora Deng Figure 4



Bonora Deng Figure 5



Bonora Deng Figure 6